PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, 15/76, 15/62, 1/21, 7/01, C07K 14/35, 16/12, A61K 39/04, 39/40, G01N 33/569, C12Q 1/04 // (C12N 1/21, C12R 1:32)

A1

(11) International Publication Number: **WO 97/08322**

(43) International Publication Date: 6 March 1997 (06.03.97)

(21) International Application Number: PCT/GB96/02015

(22) International Filing Date: 19 August 1996 (19.08.96)

(30) Priority Data:
9517494.2 25 August 1995 (25.08.95) GB

(71) Applicant *(for all designated States except US)*: THE MINISTER OF AGRICULTURE FISHERIES AND FOOD IN HER BRITANNIC MAJESTY'S GOVERNMENT OF THE UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND [GB/GB]; Whitehall Place, London SW1A 2HH (GB).

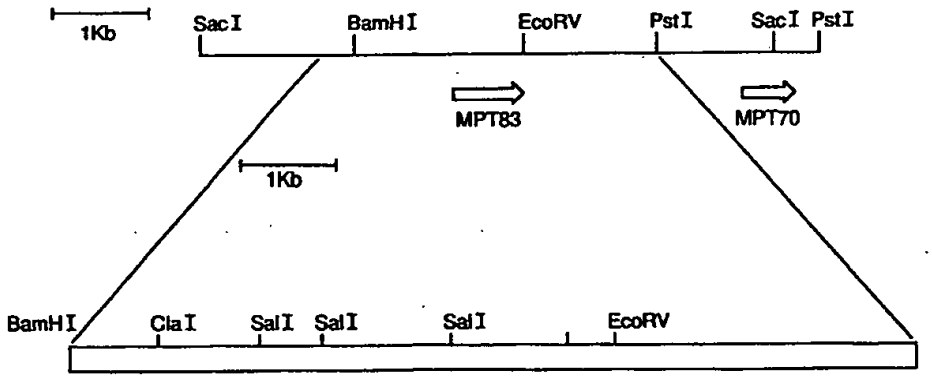
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HEWINSON, Robert, Glynn [GB/GB]; Central Veterinary Laboratory, New Haw, Addlestone, Surrey KT15 3NB (GB). JACOBS, William, Robert [US/US]; 163 Fordham Street, City Island, NY 10464 (US).

(74) Agent: GREAVES, Carol; D/IPR, Formalities Section, Poplar 2, MOD Abbey Wood #19, P.O Box 702, Bristol, BS12 7DU (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: MPT83 GENE FROM MYCOBACTERIUM TUBERCULOSIS



(57) Abstract

The invention relates to the cloning and characterisation of a novel gene from *M tuberculosis, mpt83* (Sequence ID No 2) encoding the protein MPT83 (Sequence ID No 1). Also disclosed are various substituent parts of the protein and gene including the glycosylation sequence, lipoylation sequence, secretion sequence and promoter region, and analogues and derivatives (generally 70 % homology) of these. The gene and its various aspects have utility in generating recombinant DNA expression systems for use in the transformation of cells e.g. to produce glycosylated or lipoylated products, or which are regulatable e.g. with macrophage factors. The invention also relates to recombinant antigens employing MPT83 or using the *mpt83* promoter and vaccines based on the above systems.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

MPT83 gene from Mycobacterium tuberculosis

TECHNICAL FIELD

The invention relates to the cloning and characterisation of a novel gene encoding a protein expressed in *M tuberculosis* and its analogues. The invention further relates to the glycosylation sequence, lipoylation sequence, secretion sequence and promoter region of the gene. The invention also relates generally to derived recombinant DNA expression systems and cells transformed thereby, and also recombinant antigens and applications therefor.

BACKGROUND ART

The mycobacteria are a subgroup of the actinomycetes. The include a number of organisms capable of causing serious infection in mammals such as man. Notable pathogenic agents belonging to this group are *Mycobacterium tuberculosis* and *Mycobacterium bovis*.

Identification of novel antigens of these organisms allow improved reagents for diagnostic tests or for incorporation into subunit vaccines. Additionally such antigens can be manipulated to use as serological markers.

It is known that MPB70 is an immunodominant antigen of *M bovis* and is able to elicit a delayed-type hypersensitivity response and to stimulate T lymphocyte proliferation and antibody production. MPB70 and its homologue in *M tuberculosis*, MPB70 have been cloned and sequenced and the sequences of the two genes are identical.

Although protein glycosylation is not normally found in bacteria, Fifis *et al*, in the paper "Purification and characterisation of major antigens from a *M bovis* culture filtrate", Infections and Immunity 59, 800-807, 1991 isolated three forms of MPB70 from *M bovis* which were recognised by a number of monoclonal antibodies specific for epitopes of the *M bovis* antigen MPB70. These antigens comprised a major 22kDa antigen occurred in two forms and a 25kDa antigen which was shown to be glycosylated. It was also demonstrated that carbohydrate was associated with *M bovis* antigens of 38 kDa, 25 kDa, 22 kDa and 19 kDa. More recently a

peptide from a 45 kDa antigen in *M tuberculosis* was also found to be O-glycosylated through a threonine residue(Dobos et al (1995) Infection and Immunity 63:2863-53).

The 22 kDa and a glycosylated 25 kDa antigens were reported by Fifis et al to be the non-glycosylated and glycosylated forms of the same protein and encoded by one gene as they were both recognised by a number of monoclonal antibodies specific for separate epitopes of the *M bovis* antigen MPB70.

DISCLOSURE OF THE INVENTION

The inventors have now developed a monoclonal antibody specific for the glycosylated 25 kDa antigen and have used this to establish that separate genes are responsible for production of the 22 kDa and glycosylated 25kDa antigens in *M tuberculosis*. They have determined a new gene which is distinct from the known *mpt*70 gene and which encodes the glycosylated 25 kDa antigen. The inventors have determined that homologues of these genes are present in the related species *M bovis*. Throughout this specification the *M tuberculosis* gene encoding the 25 kDa glycosylated antigen, and the homologous *M bovis* gene, will be referred to as *mpt*83 and *mpb*83 respectively.

As there is considerable homology between the *mpt*83 gene (in *M tuberculosis*) and the *mpb*83 gene (in *M bovis*), the person skilled in the art would appreciate that similar glycosylation and/or lipoylation of proteins, peptides or other products by recombinant techniques may be performed by genetic manipulation of the *mpb* 83 or any other homologous genes in related species.

The invention makes use of the *mpt*83 gene in a variety of ways. The *mpt*83 gene has been further characterised such that the promoter, glycosylation, lipoylation and secretion signals contained therein can be used to provide proteins (in particular antigens), peptides or other products, with a similar glycosylation and/or lipoylation to that determined by the *mpt*83 gene and to assist secretion from host cells recognising the secretion signal. For instance the *mpt*83 secretion signal may be

used according to a further aspect of the invention in assisting secretion of recombinant products from host cells by operatively splicing the secretion signal to the nucleotide sequence of the protein, peptide or product to be secreted from a transformed host organism recognising the secretion signal.

The proteins, peptides and products may include antigens which can subsequently be used in the development of vaccines with improved immunogenic response.

Thus the invention discloses not only the gene *mpt*83, but provides also the *mpt*83 promoter or a derivative thereof to express any protein in particular in species of bacteria, particularly Mycobacteria. Also the use of the *mpt*83 gene or derivatives thereof to glycosylate and/or lipoylate and/or secrete peptides, proteins or other products in Mycobacteria or indeed any eukaryote or prokaryote. Methods to manipulate genes to provide the above are well known in the art.

The present invention extends to any and all single or multiple nucleotide additions, deletions and/or substitutions to the gene, its glycosylation signal and/or the lipoylation and/or secretion signal and promoter and additionally to homologues of the gene, its glycosylation signal and/or the lipoylation and/or secretion signal and/or its promoter having being at least 70% identity with the authentic sequences contained herein. The present invention also extends to natural recombinant and synthetic forms of the *mpt*83 gene, its glycosylation, lipoylation and secretion signals and promoter. All such variations are referred to herein as "derivatives".

Homologues of the gene may conveniently be identified and isolated, for instance in actinomycetes or related organisms, by those skilled in the art from a test sample as follows. The test sample is contacted with a probe based on the *mpt*83 gene under suitable hybridisation conditions, and any test DNA which hybridises thereto is identified.

Such screening is initially carried out under low-stringency conditions, which comprise a temperature of about 37°C or less, a

formamide concentration of less than about 50%, and a moderate to low salt (e.g. Standard Saline Citrate ('SSC') = 0.15 M sodium chloride; 0.15 M sodium citrate; pH 7) concentration. Alternatively, a temperature of about 50°C or less and a high salt (e.g. 'SSPE'= 0.180 mM sodium chloride; 9 mM disodium hydrogen phosphate; 9 mM sodium dihydrogen phosphate; 1 mM sodium EDTA; pH 7.4). Preferably the screening is carried out at about 37°C, a formamide concentration of about 20%, and a salt concentration of about 5 X SSC, or a temperature of about 50°C and a salt concentration of about 2 X SSPE. These conditions will allow the identification of sequences which have a substantial degree of similarity with the probe sequence, without requiring the perfect homology for the identification of a stable hybrid. The phrase 'substantial similarity' refers to sequences which share at least 50% overall sequence identity. Preferably, hybridisation conditions will be selected which allow the identification of sequences having at least 70% sequence identity with the probe, while discriminating against sequences which have a lower level of sequence identity with respect to the probe.

After low stringency hybridisation has been used to identify several target sequences having a substantial degree of similarity with the probe sequence, this subset of nucleotides is then subjected to high stringency hybridisation, so as to identify those clones having a particularly high level of homology with respect to the probe sequences. High stringency conditions comprise a temperature of about 42°C or less, a formamide concentration of less than about 20%, and a low salt (SSC) concentration. Alternatively they may comprise a temperature of about 65°C or less, and a low salt (SSPE) concentration. Preferred conditions for such screening comprise a temperature of about 42°C, a formamide concentration of about 20%, and a salt concentration of about 2 X SSC, or a temperature of about 65°C, and a salt concentration of about 0.2 SSPE.

Specifically the invention provides the gene (*mpt*83) which encodes the glycosylated 25 kDa antigen of *Mycobacterium tuberculosis* or a derivative thereof.

The invention also provides the promoter of the *mpt*83 gene or a derivative thereof; the DNA sequence encoding the glycosylation signal and/or the lipoylation and/or secretion signal of the *mpt*83 gene or derivatives thereof; DNA sequences encoding the glycosylation signal and/or the lipoylation and/or secretion signal of the *mpt*83 gene or derivatives thereof. When using a promoter sequence the invention provides for a DNA molecule comprising a promoter according to the invention operably linked to a heterologous protein coding sequence; a DNA molecule comprising a nucleotide sequence encoding the glycosylation signal and/or the lipoylation and/or secretion signal according to the invention operably linked to a heterologous protein coding sequence.

The present invention also therefore extends to DNA constructs comprising portions or derivatives of the *mpt*83 gene and/or its glycosylation sequence and/or its lipoylation sequence and/or its secretion sequence and/or its promoter region attached to a heterologous nucleotide sequence operatively linked in a 5' to 3' direction.

The present invention also extends to organisms transformed with such DNA constructs and particularly species of actinomycetes.

Known techniques enable functional portions of genes to be spliced to nucleotides sequences. In this way the DNA encoding the glycosylation signal from the *mpt*83 gene can be spliced to a gene sequence encoding a peptide protein or other product such that expression of the spliced gene causes the protein to be glycosylated, provided the host cell has the machinery for recognising such a glycosylation signal and applying the glycosylation.

Similarly splicing the secretion signal to a gene sequence encoding a protein peptide or other product allows secretion thereof in a recombinant host cell organism. The heterologous protein coding sequence may encode an antigenic protein of a pathogen. This has particular utility in the application of vaccines. This may be performed by forming a suitable vector with

which to transform a host production organism which can be a prokaryote or eukaryote cell.

Optionally the promoter sequence of the *mpt83* gene may be ligated such that any protein can be glycosylated and/or lipoylated and expressed in mycobacteria, other actinomycetes recognising the promoter. Alternatively the spliced nucleotide sequence of a protein peptide or other product and the *mpt*83 glycosylation signal and/or its lipoylation and/or the secretion signal may be used with a suitable vector or otherwise to transform any prokaryote or eukaryote cells. The use of recombinant DNA vectors as tools in such expression systems is well known. The invention covers suitable transformation tools such as vectors containing the *mpt*83 gene sequence, and/or the promoter thereof and or the glycosylation and or lipoylation signal thereof and/or the secretion sequence thereof. Such genes, coding sequences, promoters, DNA sequences encoding the glycosylation signal and/or the lipoylation and/or secretion signals, DNA molecules and vectors may be obtained in isolated form and in substantially pure form.

The invention further extends to prokaryote or eukaryote cells transformed the above mentioned DNA constructs, e.g. transformed with vectors or by other means e.g. electroporation. This enables any peptide, protein or other product to be glycosylated, lipoylated. Using the secretion signal enables the recombinant product to be excreted outside the host cell when recombinant product produced in a suitable host.

The invention covers any host cell but in particular, for example an actinomycete such as a mycobacterium such as *M tuberculosis* or *M bovis* (e.g. *M bovis* BCG), transformed or transfected with such DNA constructs or vectors according to the invention. The advantages of using mycobacteria is that the glycosylation and lipoylation machinery is likely to be available and the secretion signal recognised by other cellular components. Such host cells may have utility as vaccine organisms.

The manipulation of the *mpt*83 gene in a vaccine organism may provide a serological marker, for instance allowing differentiation between vaccine and wild-type organisms e.g. pathogenic organisms, using e.g. monoclonal antibodies. Such manipulation may involve preventing expression of the *mpt*83 gene product, or causing an epitope-deficient mutant to be expressed.

The invention also extends to a vaccine comprising a pharmaceutically acceptable carrier or diluent and, as active ingredient, a vaccine organism according to the invention.

The invention is also particularly useful in providing recombinant antigens wherein the protein peptide or other product referred to above is an antigen i.e. from a pathogenic organism. The invention also extends to a vaccine comprising a pharmaceutically acceptable carrier or diluent and such an antigen or an epitopic part thereof.

Also a method for vaccinating a patient against infection by a pathogen, which method comprises administering to the patient a host cell transformed according to the invention with *mpt*83 DNA sequences; or a protein comprising the following elements linked in a N-terminal to a C-terminal direction: the glycosylation signal and/or lipoylation of the MPT83 or a derivative thereof and a heterologous antigenic sequence of a pathogen.

The invention will now be explained with reference to the following non-limiting Figures and Examples. Further embodiments falling within the scope of the invention will occur to those skilled in the art in the light of these.

SEQUENCE LISTINGS

Sequence ID No 1 is the nucleotide and amino acid sequence of the *mpt*83 gene product MPB83.

Sequence ID No 2 is the *mpt83* sequence cloned into pSM13, including the promoter region.

FIGURES

Figure 1 is a restriction map of the *mpt*83 region of the *M tuberculosis* genome. *Eco*RV, *Bam*HI, *Cla*I, *Pst*I, *Sac*I and *Sal*I restriction endonucleases sites are shown with a scale of 1Kb (kilobases), and also a smaller scale of the *mpt*83 region which was cloned into the plasmid pSM13.

Figure 2a shows a large portion of the *mpt83* promoter sequence that was produced during the characterisation of the full gene. The first *Cla*I site (bases 10-15 on the Figure) corresponds to the *Cla*I site at bases 400-405 on Sequence ID No 2.

Figure 2b shows a large portion of the *mpt83* gene coding region that was produced during the characterisation of the full gene. The first base ('G') corresponds to base number 171 on Sequence ID No 1.

EXAMPLES

EXAMPLE 1: A DESCRIPTION OF THE MPT83 GENE

The *mpt*83 gene is located generally within the region defined by the BamH1 and Pst1 restriction sites of Figure 1, or alternatively the region defined by the (same) *Bam*H1 site and the *Eco*RV sites and may be defined as the open reading frame of these sequences.

It may be noted that the first *Sal*I restriction site marked in Figure 1 and Figure 2a was putatively identified as such on the basis that it was cut by *Hinc*II, which is specific *inter alia* for the *Sal*I sequence. The original sequence has since been revised slightly and this site is marked as *Hinc*II on Sequence ID No 2.

The *mpt*83 region of the *M tuberculosis* genome may therefore be obtained in any suitable manner, such as excision using the above mention restriction enzymes. For example, total genomic DNA may be extracted from *M tuberculosis* and digested to completion using *Bam*HI and *Pst*I. The resultant DNA is fractionated. DNA fragments of an appropriate size, for example 2 to 3 Kb, are ligated into a vector. The ligation mix is used for transformation or

electroporation of competent recipient *E Coli*. Recombinant colonies are screened for the presence of the *mpt*83 region. More detail is given in Example 2 below.

Sequence ID No 2 shows the complete sequence of the *mpt*83 gene. The coding sequence of the gene begins at position 1149 (this is position 1 on Sequence ID No 1). Sequences upstream to this comprise the promoter region which initiates transcription of the gene. A typical TATAA box is present between positions 1065-1069. A selection of restriction endonuclease sites are listed as features in Sequence ID No 2. Sequence analysis of the gene shows a 660 bp open reading frame encoding protein of 220 amino acids with a predicted molecular weight of 22 kDa, and a predicted isoelectric point of 4.7. A Shine-Delgarno sequence (AAGGAA) is present 9 nucleotides upstream of the initiation (ATG). The N-terminal protein *mpt*83 revealed a typical prokaryotic secretion signal peptide and the presence of a lipoprotein signal peptide cleavage motif between amino acid residues 22-26, Sequence ID No 1.

There is 65% homology with *mpt*70. The fact that it contains a typical prokaryote lipoprotein signal peptide signal cleavage motif suggests that *mpt*83 is a lipoprotein anchored in the cell wall by a thioester linkage of the amino-terminal cysteine residue to the fatty acid moiety of the cell membrane. The 2kDa size difference between *mpt*83 expressed in *E coli* and in both *M smegmatis* and *M bovis* implies that the protein undergoes considerable posttranslational modification in mycobacteria. An amino acid sequence comparison between *mpt*83 and *mpt*70 reveals the presence of a proline rich peptide sequence upstream of two threonines and two alanines in *mpt*83 which is not present in *mpt*70. The gene contains glycosylation, lipoylation and secretion sequences which are detailed below.

EXAMPLE 2: DETAILED ISOLATION OF THE *MPT*83 REGION OF THE *M TUBERCULOSIS*; TRANSFORMATION AND EXPRESSION USING THE GENE

METHOD 1: To isolate the *mpt83* region, total genomic DNA is extracted from *M tuberculosis* using standard protocols e.g. Z

Zainuddin Dale JW Journal of General Microbiology 135, pp2347 to 2355, 1989. 10 micrograms of DNA are digested to completion using BamH1 and Pst1 jointly (100 units of each) in commercially available universal buffer. The resultant DNA is fractionated by agarose gel electrophoresis with TAE buffer and 0.8% low gelling temperature agarose. The gel is stained with ethidium bromide. The DNA is visualised under UV irradiation. An agarose block containing DNA fragments in the region 2 to 3 Kb in size as deduced by molecular weight markers is excised and the DNA purified by commercially available glass milk systems. Vector DNA e.g. pUC18 (1 microgram) is digested to completion with BamH1 and Pst1 (10 units of each) in the universal buffer. The cut vector is purified by PAGE and glass milk. Purified vector and insert DNA are quantified separately either by absorbent 280/260 nm wavelength or by a commercially available dipstick. Equimolar amounts of vector and insert DNA are mixed to a volume not exceeding 15 microlitres to which is added ligase buffer (5 microliters x 5 conc.) and 1 microlitre T4 DNA ligase (2.5 units). DNA is incubated at 40°C for 16 hours 2.5 microlitres of ligation mix is used in standard transformation or electroporation of competent recipient *E Coli*. Selection can be made for recombinants on LB agar containing ampicillin 50 micrograms per ml with standard blue-white selection using X-dal and IPTG. White recombinant colonies are screened for the presence of MPT83 by hybridisation or immunoscreening with monoclonal MBS43 specific for the gene product (Goodger *et al*, Vet. Record 135, 82-85, 1994).

METHOD 2: An alternative is to clone the BamH1-EcoRv fragment. This may be obtained by suitable restriction enzyme cutting according to the manufacturers instructions. Bacteria such as *E coli* strains DH5 and HB101 and *M smegmatis* strain mc^2-155 can be transformed with the *mpt*83 gene using a suitable vector. For example a cosmid library may be constructed of *M tuberculosis* genomic DNA in a suitable plasmid/cosmid such as pYUB18. Transformation may be performed by any suitable method in the art as above or by electroporation.

Expression of the *mpt*83 protein may also be ascertained by immunoblotting of *M bovis* or *M tuberculosis* lysates produced by e.g. sonication with any suitable specific monoclonal antibody. The monoclonal antibody MBS43, for example, recognises the 25kDa antigen i.e. the MPT83 antigen. Immunoblotting of the culture filtrate and cell lysate shows that the 25kDa antigen is predominantly associated with the cell pellet; the *mpt*70 antigen is predominant secreted.

EXAMPLE 3: PREPARATION OF *MPT*83 PROMOTER BY PCR

To amplify the *mpt*83 promoter region, specific primers which are 5'GGATCCGTGGTAGGGGATGT-3' (positions 1 to 20, Sequence ID No 2) and 5'CGCTTCGATTCCTTTGCTTC-3' (complementary positions 1148 to 1128, Sequence ID No 2) are used with DNA extracted as described earlier from *M tuberculosis*. The procedure is essentially that described by Saiki *et al* (1988) Science 239:487. The buffer is titrated to contain the optimum level of magnesium ions and the thermal cycling conditions are as follows: the DNA is denatured at 98°C for 5 minutes and followed by 40 cycles at 94°C for 1 minute 62°C for 1 and a half minutes for annealing and two minutes at 72°C for extension. A *Taq* polymerase is used for amplification of the promoter region to give a product with a predicted size of 1148 bp. To identify the product, samples of the PCR are run on an agarose gel and Southern blots are hybridised with the SalI fragment (positions 641 to 832, Sequence ID No 2) labelled with any reporter molecule.

The full-length *mpt*83 promoter is defined within the sequence from position 1 to position 1148 shown in Sequence ID No 2. A part of this sequence may however be capable of acting as a promoter and thus forms part of the invention. The full-length or part-length promoter sequence may be obtained by using the polymerase chain reaction (PCR) or may be synthesised. A part-length promoter sequence therefore may be obtained by removing nucleotides from the 5'-terminus or the 3'-terminus of the full-length promoter sequence, for example using an exonuclease such as exonuclease III or BAL 31. A modified promoter sequence may be obtained by introducing changes into the full-length or part-length promoter

sequence, for example substitutions, insertions and/or extensions. This may be achieved by any appropriate technique, including restriction of the full-length or part-length sequence with an endonuclease, insertion of linkers, use of an exonuclease and/or a polymerase and site-directed mutagenesis. A promoter according to the invention may have a sequence at least 70% identical to the sequence of the full-length or part-length promoter. There may be, for example up to ten or up to twenty nucleotide deletions, insertions and/or substitutions made to the full-length or part-length promoter sequence. Whether a part-length or modified sequence is capable of acting as a promoter may be readily ascertained. The candidate promoter sequence is provided in a vector upstream of a protein coding sequence e.g. a marker promoter. It is provided at a location at which it is believed to be operatively linked to the protein coding sequence. A suitable host is transformed with the resulting vector. The presence or absence of expression of the encoded protein is then determined.

As described in Example 6, a DNA vector comprising the promoter of *mpt*83 gene or its derivatives may be used to transform a host cell so as to direct the transcription of nucleotide sequences operatively linked downstream thereof. Someone skilled in the art would readily realise that to facilitate an insertion of a nucleotide sequence, encoding a peptide protein or other product, into such a vector would require one or more unique restriction sites down steam of the promoter, so that the vector can be cleaved at such sites and the sequence inserted. The nucleotide sequence may well be the *mpt*83 gene itself. The vectors can subsequently be utilised to transform actinomycetes or other related organisms such as mycobacteria species (e.g. *M tuberculosis* or *M bovis*), and then use such a transformed organism for production of the peptide, protein or product encoded by the nucleotide sequence.

EXAMPLE 4: THE GLYCOSYLATION AND LIPOYLATION SEQUENCES OF MPT83

The full-length *mpt*83 glycosylation sequence is defined by the nucleotide sequence from positions 1251-1319, Sequence ID No 3.

The glycosylation signal as represented by the 23 amino acid sequence is defined as:

TSPKPATSPAAPVTTAAMADPAA

These are amino acids 35-57 on Sequence ID No 1.

To amplify the *mpt*83 glycosylation region, specific primers which are 5'-ACCAGCCCGAAACCGGCGA-3' (positions 1251 to 1269, Sequence ID No 2) and 5'-CGCTGCGGGGTCAGCCATT-3' (complementary positions 1319 to 1301, Sequence ID No 2) are used with DNA extracted as described earlier from *M tuberculosis*. The procedure is essentially as described in Example 2. To identify the product, samples of the PCR products are run on an agarose gel and Southern blots are hybridised with a probe having a nucleotide sequence corresponding to bases within the primer-defined area e.g. 1270-1301 of Figure 2b.

The full-length *mpt*83 leader plus lipoylation sequence is defined by the nucleotide sequence from positions 1149-1244, Sequence ID No 2. The secretion/lipoylation signal as represented by amino acid sequence is defined as :

MINVQAKPAAAASLAAIAIAFLAGCSSTKPVS

This is amino acids positions 1-32, Sequence ID No 1. The lipoylation motif itself is:

LAGCS

This is amino acids positions 22-26, Sequence ID No 1

Modified glycosylation or lipoylation sequences may be obtained by introducing changes into the full-length or part-length sequences respectively, for example substitutions, insertions and/or extensions. This may be achieved by any appropriate technique, including restriction of the full-length sequence with an endonuclease, insertion of linkers, use of an exonuclease and/or a polymerase and site-directed mutagenesis. A glycosylation or lipoylation sequence according to the invention may be at least 70% identical to the sequence of the full-length or part-length

glycosylation sequence. There may be up to five, for example up to ten or up to twenty nucleotide deletions, insertions and/or substitutions made to the full-length or part-length promoter sequence. Part of these full-length sequences may however be capable of directing glycosylation lipoylation respectively and thus form part of the invention. The full-length or part-length glycosylation or lipoylation sequence may be obtained by using the polymerase chain reaction (PCR) or may be synthesised. A part-length glycosylation or lipoylation sequence therefore may be obtained by removing nucleotides from the 5'-terminus or the 3'-terminus of the full-length sequence, for example using an exonuclease such as exonuclease III or BAL 31.

Use can be made of the glycosylation sequence or lipoylation sequence of *mpt* 83 (or their derivatives) to respectively glycosylate or lipoylate any protein, peptide or other product from any eukaryote or prokaryote by ligating these sequences to nucleotide sequences encoding said peptide, protein or other product and transformation of suitable host organisms therewith. Recombinant DNA vectors, for instance as described in Example 6, are useful tools in this procedure. In essence such a vector is constructed wherein the glycosylation or lipoylation signal sequence from *mpt*83 is operably linked to a nucleotide sequence encoding the peptide, protein or product to be glycosylated or lipoylated as appropriate. In addition such a vector can include restriction endonuclease sites such that transcription and translation of a nucleotide sequence inserted into said sites results in a glycosylated or lipoylated transcription and translation product of the nucleotide sequence. Whether a part-length or modified sequence is capable of causing glycosylation or lipoylated of an expressed polypeptide may be readily ascertained. A promoter operatively linked to the protein coding sequence may also be provided and may include the *mpt*83 promoter or a derivative thereof. The candidate glycosylation or lipoylation sequence is provided at a suitable location with respect to the protein coding sequence. A suitable host is transformed with the resulting vector. The presence or absence of glycosylation or lipoylation of the expressed protein is then determined.

EXAMPLE 5: USE OF THE SECRETION SIGNAL OF MPT83

The full-length *mpt*83 secretion signal glycosylation sequence is defined by base positions 1149 to 1211, Figure 2b . Part of these full-length sequences may however be capable of directing secretion and thus form part of the invention. The full-length or part-length secretion sequence may be obtained by using the polymerase chain reaction (PCR) or may be synthesised.

The secretion signal as represented by amino acid sequence is defined as :

MINVQAKPAAAASLAAIAIAF

This is amino acids position 1-21, Sequence ID No 1.

Modified secretion sequence may be obtained by introducing changes into the full-length or part-length sequences respectively, for example substitutions, insertions and/or extensions. A secretion sequence according to the invention may be at least 70% identical to the sequence of the full-length or part-length glycosylation sequence

Use can be made of the secretion signal in recombinant systems whereby the secretion signal is ligated so as to be operably linked to any nucleotide sequence encoding a protein peptide or other product. Transformation of organisms and in particular mycobacteria allow the protein peptide or other product to include the secretion signal so that the transcribed peptide protein or other product is secreted from the bacterial cell wall.

EXAMPLE 6: DNA CONSTRUCTS

The invention includes DNA constructs such as vectors comprising the nucleotide sequence of the *mpt*83 promoter sequence and/or glycosylation sequence and/or the lipoylation sequence and/or the secretion sequence (or derivatives of these) operatively linked either to the nucleotide sequence of the *mpt*83 translation product or the nucleotide sequence encoding of any protein, peptide of product of any prokaryote or eukaryote.

Recombinant DNA vectors/constructs such as plasmids are therefore useful in transforming organisms particularly mycobacteria such that they can be made to express any protein peptide or other product. If one requires to glycosylate and or lipoylate a peptides, protein or other product from any other product the vector would include the nucleotide sequence therefore operatively linked to the glycosylation sequence and /or lipoylation sequence. In essence such a vector is constructed wherein the glycosylation or lipoylation signal sequence from *mpt*83 is operably linked to a nucleotide sequence encoding the peptide, protein or product to be glycosylated or lipoylated as appropriate. In addition such a vector/DNA construct can include restriction endonuclease sites that allow transcription and translation of a nucleotide sequence inserted into said sites such that a glycosylated or lipoylated translation product of the nucleotide sequence is produced. Whether a part-length or modified sequence is capable of causing glycosylation or lipoylated of an expressed polypeptide may be readily ascertained. A promoter operatively linked to the protein coding sequence may also be provided; this may be the *mpt*83 promoter or a derivative thereof The candidate glycosylation or lipoylation sequence is provided at a suitable location with respect to the protein coding sequence. A suitable host can be transformed with the resulting vector. The recombinant sequence may be inserted behind a strong promoter of the host organism to enhance expression. The presence or absence of glycosylation or lipoylation of the expressed protein is then determined to ascertain that successful transformation has occurred.

The addition of the secretion signal to the vector enables the recombinant product to be secreted from the cell wall when expressed in a suitable host such as mycobacteria species.

Alternatively, other mycobacterial promoters or promoters effective in actinomycetes and other species including mycobacteria could be cloned into the *mpt*83 gene or fusion vector systems containing segments of the *mpt*83 gene as described such as to initiate the transcription of the sequences of the *mpt*83 gene which may facilitate both translation and secretion when cloned in the *mpt*83 fusion vector.

As used in the present specification, a "DNA vector" is defined to include plasmid DNA, lysogenic phage DNA and/or transposon DNA, in double or single stranded linear or double or single stranded circular form and may or may not be self transmissible or mobilizable. A person skilled in the art will realise that one or more origins of replication may be required. At least one origin of replication will allow replication in a prokaryotic cell and preferably in an actinomycetes or related organism and more preferably in a species of mycobacterium. The vector may be introduced into a cell by any of a number of techniques such as conjugation, mobilization, transformation, transfection, transduction or electroporation amongst others. Transformation of eukaryotic or prokaryotic cells with the vectors and DNA is by introduction of the vector or DNA by any means and is either integrated into the cells genome or existing extrachromosomally (e.g. as an autonomously replicating plasmid). Linker sites comprising one or more restriction enzyme sites can be introduced into the gene at any point after the promoter region, by site directed mutagenesis, partial cleavage and linker mutagenesis, transposition of restriction sites or other techniques.

Vectors may be any indigenous plasmid, lysogenic bacteriophage or transposon from any of these bacteria which may be used as a source of the DNA vector including shuttle vectors. By a shuttle vector as used herein is meant a DNA vehicle capable, naturally or by design, of replication in both the actinomycetes and related species and in bacteria such as E coli, Bacillus sp. or Pseudomonas sp. and the like and/or eukaryotic cells such as mammalian, yeast or fungal cells. In this case, the vectors contemplated herein will contain at least one origin of replication wherein where necessary replication can occur in a prokaryotic cell and further origins of replication for replication in different prokaryotic species and/or eukaryotic species. Additionally, although the present invention is useful for any actinomycete or related organisms, it is particularly directed to any species of *M bovis* BCG.

Although the DNA constructs of the present invention are described in terms of DNA, it would be readily apparent to one skilled in

the art that the constructs could be maintained and in some cases used in corresponding RNA form without falling outside the scope of the present invention.

The invention also includes methods of producing recombinant products such as proteins peptides or other products by transforming any suitable host organism with vectors as describe hereinbefore and the recombinant products formed therefrom.

EXAMPLE 7: THE USE OF MACROPHAGES IN EXPRESSION OF THE MPT83 ANTIGENS

The search to date for antigens of *M bovis* and *M tuberculosis* for use in immunodiagnosis and vaccination has relied on the identification of antigens that are expressed when the organism is grown on synthetic medium: it is known that virulence determinants in some bacterial species may be controlled by environmentally responsive regulators, such that they react to phagocytosis by macrophages with increased expression of macrophage-induced proteins (MIPs). The inventors have determined that the *mpt*83 gene is up-regulated in macrophages, that is to say infecting macrophages with *M bovis* BCG strain Pasteur results in enhanced production of the glycosylated product. Culturing transformed mycobacteria of the MTb complex with macrophages or macrophage factors can therefore be used to enhance yields of recombinant proteins peptides or other products referred to above.

The antigenic profiles obtained by Western blotting BCG grown inside macrophages against sera from a calf experimentally infected with *M bovis* reveals a number of antigenic differences including the up-regulation of antigens of the 25 kDa form of MPT83 in response to the macrophage environment. The up-regulation of the 25 kDa protein for *M tuberculosis* grown inside the macrophage cell line 1C-21 or when subjected to heat shock at 45 or 48°C has also been reported by Alavi & Affronti (1994) J Leukocyte Biology 55: 633-641. The expression reached its maximal level of expression about 7 h post infection. Thus it will be apparent to the skilled person that the *mpt83* promoter may have

utility in directing the up-regulation of expression of heterologous genes inside host macrophages.

EXAMPLE 8: VACCINES

The heterologous protein coding sequence referred to above which may be operatively linked to the *mpt*83 promoter and/or glycosylation and/or lipoylation and/or secretion signals may include the heterologous antigenic sequence of a pathogen i.e. may encode a sequence of amino acid residues capable of raising antibody to epitopic parts of the antigen. According to the invention antigenic products from such pathogens can provided with glycosylation or lipoylation according to the invention. In this way subunit vaccines against pathogens may be provided with antigens or epitopic parts thereof having glycosylation and/or lipoylation similar to *mpt*83. In order to produce such antigens mycobacteria and in particular *M smegmatis, M vaccae* and BCG are preferably used as host organisms to produce such recombinant antigens. The use of the secretion signal in combination with the antigenic sequence allows the recombinant antigens to be secreted from the cell and therefore assists production of the antigens. The recombinant antigens can then be used in vaccines.

In addition live recombinant cells incorporating antigenic coding sequences as described above may be used in vaccines to induce an immune response. The invention further therefore extends to recombinant live transformed cells incorporating antigenic coding sequences having glycosylated / lipoylated or secretion signals from MTP83. The sequence may thus encode an antigen capable of raising neutralising antibody, for example an antigen of an infectious agent or pathogen such as a virus, bacterium or parasite. As examples of viruses whose antigens may be presented there may be mentioned hepatitis A virus, hepatitis B virus, hepatitis C virus, foot-and-mouth disease virus, poliovirus, herpes simplex virus, rabies virus, human immunodeficiency virus type 1, (HIV-1), HIV-2, simian immunodeficiency virus (SIV), human rhinovirus (HRV), dengue virus and yellow fever virus. The antigen of a parasite may be a malarial antigen, such as an antigen of *Plasmodium falciparum.*

For vaccination against *M tuberculosis* itself and *M bovis* or related species the active component of a vaccine therefore may include so called "naked DNA". This therefore covers the use of portions of the *mpt*83 gene (or derivative of) for use as vaccine components.

The invention extends to vaccines including recombinant antigens as described above, or vaccines containing live cells transformed with *mpt*83 nucleotide sequences. They may be administered in any appropriate fashion. Thus, they may be administered orally or parenterally, for example intradermally, intravenously or intramuscularly as appropriate. They may be formulated with appropriate carriers or diluents such as physiological saline. An adjuvant such as aluminium oxide may be provided. Appropriate dosages may be selected depending upon the particular vaccine to be administered and the host to which the vaccine is being given. It is particularly preferred to administer in this way *M bovis* BCG transformed according to the invention.

In addition to using the *mpt*83 gene, its glycosylation and /or its lipoylation and/or its secretion signal and/or its promoter region, or combinations thereof, in the construction of a variety of vector molecules, the present invention contemplates using these nucleotide sequences to delete or insert the *mpt*83 gene, or parts thereof, in *M bovis, M bovis* BCG or any other mycobacterial species for the purpose of creating a vaccine inducing a serological or cell-mediated Immune response with a pathogenic mycobacterium. This has important consequences in diagnostic research and disease management in that individuals vaccinated carrying a deletion mutation in the *mpt*83 gene, will mount a serological and cell-mediated immune response diagnostically distinguishable from an individual vaccinated with a strain expressing the *mpt*83 gene. Techniques for inserting genes or inducing insertional or deletional mutations are well known in the art.

M bovis BCG is an attractive candidate delivery vehicle for the development of new multivalent recombinant vaccines since BCG may be delivered orally is safe heat stable cheap and has well-

documented adjuvant properties. Many viral and protozoan antigens are glycosylated and it is possible that antigens requiring glycosylation to stimulate a protective response could be engineering to contain glycosylation signals recognised by BCG.

EXAMPLE 9: DIAGNOSTIC TESTING FOR MYCOBACTERIA

The invention additionally extends to the use of *mpt*83 gene products including recombinant antigens as described hereinabove or epitopic parts thereof for use in testing for infection of humans and animal with species of mycobacteria and in particular *M tuberculosis* and *M bovis*. The *mpt*83 gene products may be provided either unconjugated or conjugated with suitable reagents and used in diagnostic kits to indicate infection when reacted with e.g. blood, sera of infected animals. Such techniques are well known in the art.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: THE MINISTER FOR AGRICULTURE, FISHERIES AND FOOD ETC.
(B) STREET: WHITEHALL PLACE
(C) CITY: LONDON
(E) COUNTRY: UNITED KINGDOM
(F) POSTAL CODE (ZIP): SW1A 2HH

(A) NAME: ROBERT GLYNN HEWINSON
(B) STREET: CENTRAL VETERINARY LABORATORY
(C) CITY: ADDLESTONE, NEW HAW
(D) STATE: SURREY
(E) COUNTRY: UNITED KINGDOM
(F) POSTAL CODE (ZIP): KT15 3NB

(A) NAME: WILLIAM R JACOBS JR.
(B) STREET: HOWARD HUGHES MEDICAL INSTITUTE, ALBERT EINSTEIN COLLEGE
(C) CITY: BRONX
(D) STATE: NEW YORK
(E) COUNTRY: UNITED STATES OF AMERICA
(F) POSTAL CODE (ZIP): NONE

(ii) TITLE OF INVENTION: RECOMBINANT DNA EXPRESSION SYSTEM

(iii) NUMBER OF SEQUENCES: 2

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: GB 9517494.2
(B) FILING DATE: 25-AUG-1995

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 663 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:
(A) NAME/KEY: CDS

(B) LOCATION: 1..663
(D) OTHER INFORMATION: /product= "MPT83"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..96
(D) OTHER INFORMATION: /function= "Lipoylation motif"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..63
(D) OTHER INFORMATION: /function= "secretion signal"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 103..171
(D) OTHER INFORMATION: /function= "Glycosylation motif"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
ATG ATC AAC GTT CAG GCC AAA CCG GCC GCA GCA GCG AGC CTC GCA GCC        48
Met Ile Asn Val Gln Ala Lys Pro Ala Ala Ala Ala Ser Leu Ala Ala
  1               5                  10                  15

ATC GCG ATT GCG TTC TTA GCG GGT TGT TCG AGC ACC AAA CCC GTG TCG        96
Ile Ala Ile Ala Phe Leu Ala Gly Cys Ser Ser Thr Lys Pro Val Ser
            20                  25                  30

CAA GAC ACC AGC CCG AAA CCG GCG ACC AGC CCG GCG GCG CCC GTT ACC       144
Gln Asp Thr Ser Pro Lys Pro Ala Thr Ser Pro Ala Ala Pro Val Thr
        35                  40                  45

ACG GCG GCA ATG GCT GAC CCC GCA GCG GAC CTG ATT GGT CGT GGG TGC       192
Thr Ala Ala Met Ala Asp Pro Ala Ala Asp Leu Ile Gly Arg Gly Cys
    50                  55                  60

GCG CAA TAC GCG GCG CAA AAT CCC ACC GGT CCC GGA TCG GTG GCC GGA       240
Ala Gln Tyr Ala Ala Gln Asn Pro Thr Gly Pro Gly Ser Val Ala Gly
65                  70                  75                  80

ATG GCG CAA GAC CCG GTC GCT ACC GCG GCT TCC AAC AAC CCG ATG CTC       288
Met Ala Gln Asp Pro Val Ala Thr Ala Ala Ser Asn Asn Pro Met Leu
                85                  90                  95

AGT ACC CTG ACC TCG GCT CTG TCG GGC AAG CTG AAC CCG GAT GTG AAT       336
Ser Thr Leu Thr Ser Ala Leu Ser Gly Lys Leu Asn Pro Asp Val Asn
            100                 105                 110

CTG GTC GAC ACC CTC AAC GGC GGC GAG TAC ACC GTT TTC GCC CCC ACC       384
Leu Val Asp Thr Leu Asn Gly Gly Glu Tyr Thr Val Phe Ala Pro Thr
        115                 120                 125

AAC GCC GCA TTC GAC AAG CTG CCG GCG GCC ACT ATC GAT CAA CTC AAG       432
Asn Ala Ala Phe Asp Lys Leu Pro Ala Ala Thr Ile Asp Gln Leu Lys
    130                 135                 140

ACT GAC GCC AAG CTG CTC AGC AGC ATC CTG ACC TAC CAC GTG ATA GCC       480
Thr Asp Ala Lys Leu Leu Ser Ser Ile Leu Thr Tyr His Val Ile Ala
145                 150                 155                 160
```

```
GGC CAG GCG AGT CCG AGC AGG ATC GAC GGC ACC CAT CAG ACC CTG CAA        528
Gly Gln Ala Ser Pro Ser Arg Ile Asp Gly Thr His Gln Thr Leu Gln
                165                 170                 175

GGT GCC GAC CTG ACG GTG ATA GGC GCC CGC GAC GAC CTC ATG GTC AAC        576
Gly Ala Asp Leu Thr Val Ile Gly Ala Arg Asp Asp Leu Met Val Asn
            180                 185                 190

AAC GCC GGT TTG GTA TGT GGC GGA GTT CAC ACC GCC AAC GCG ACG GTG        624
Asn Ala Gly Leu Val Cys Gly Gly Val His Thr Ala Asn Ala Thr Val
        195                 200                 205

TAC ATG ATC GAT ACG GTG CTG ATG CCC CCG GCA CAG TA                     663
Tyr Met Ile Asp Thr Val Leu Met Pro Pro Ala Gln
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1898 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION: 1..1148

(ix) FEATURE:
(A) NAME/KEY: misc_signal
(B) LOCATION: 1065..1069
(D) OTHER INFORMATION: /standard_name= "TATAA BOX"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1135..1140
(D) OTHER INFORMATION: /standard_name= "SHINE-DELGARNO SEQUENCE"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 400..405
(D) OTHER INFORMATION: /standard_name= "ClaI site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 641..646
(D) OTHER INFORMATION: /standard_name= "HincII site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 833..838
(D) OTHER INFORMATION: /standard_name= "SalI site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1488..1493

(D) OTHER INFORMATION: /standard_name= "SalI site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1892..1898
(D) OTHER INFORMATION: /standard_name= "EcorV site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GGATCCGTGG TAGGGGATGT CGCCGACCCG CTGCGCCGCG TGCTCGTCAG CGACGGCAAT 60

ATTGGTCACC CCGGTCTGCG CACGTTGTCG CAGCAACGTG TCCAGATGGG CGCCGCCAGC 120

GGCCAGCCCG ACTACCTCGA AACGGTCCGG ATTGTCGGCG ATGACCTGAA GCGCCTGGGT 180

GCCGATCGAG CCGGTACTGC CCAGCACCAC CACCCGCAAC CGGCCGTCAG CGCGCCCGTC 240

GGTCGAGTTG GTCACCTCAT CATTGTGCGC CACCACCTCG TTGTCACCGC GCCGCCGGAT 300

CACGACGCGT CCACCGGTAG CCACACTTCC CCGTGGAATG CAATCGTCTT GATGCCTGCG 360

CTTGATGCTA AGATGCCATG CGTGCGCACG ACGATCCGTA TCGATGACGA GCTGTACCGC 420

GAGGTGAAAG CAAAGGCCGC TCGTTCCGGG CGTACCGTGG CCGCGGTTCT TGAAGATGCG 480

GTGCGGCGTG GTCTCAACCC GCCTAAGCCG CAGGCCGCCG GCCGTTATCG AGTCCAGCCG 540

TCGGGTAAGG GCGGCCTGCG GCCCGGTGTC GATCTATCGT CCAACGCCGC ACTTGCCGAA 600

GCGATGAACG ACGGCGTGTC GGTCGATGCT GTGCGTTGAT GTCAACGTGC TCGTTTACGC 660

GCATCGGGCA GACCTACGGG AGCACGCGGA CTATCGGGGT TTGCTTGAGC GGCTGGCCAA 720

CGATGACGAG CCGCTGGGTC TACCAGATAG CGTGCTCGCC GGCTTCATCC GGGTGGTTAC 780

CAACCGCCGC GTCTTCACCG AGCCGACGAG CCCACAGGAC GCATGGCAGG CAGTCGACGC 840

CCTACTCGCG GCACCCGCAG CCATGCGACT TCGGCCTGGC GAGCGCCACT GGATGGCCTT 900

TCGGCAGTTA GCGTCCGATG TTGATGCGAA CGGCAACGAC ATTGCGGACG CGCACCTGGC 960

CGCCTACGCG CTAGAGAACA ACGCAACCTG GTTGAGCGCC GACCGCGGCT TTGCCCGTTT 1020

CCGTCGACTG CGCTGGCGTC ATCCGTTGGA CGGTCAGACC CATCTATAAC CGGCCCCACT 1080

CCGAATCACT GGTGTCCACC CAGGAGGACG GCGTTCAACG CCGCCGCAGA AGCAAAGGAA 1140

TCGAAGCGAT GATCAACGTT CAGGCCAAAC CGGCCGCAGC AGCGAGCCTC GCAGCCATCG 1200

CGATTGCGTT CTTAGCGGGT TGTTCGAGCA CCAAACCCGT GTCGCAAGAC ACCAGCCCGA 1260

AACCGGCGAC CAGCCCGGCG GCGCCCGTTA CCACGGCGGC AATGGCTGAC CCCGCAGCGG 1320

ACCTGATTGG TCGTGGGTGC GCGCAATACG CGGCGCAAAA TCCCACCGGT CCCGGATCGG 1380

TGGCCGGAAT GGCGCAAGAC CCGGTCGCTA CCGCGGCTTC CAACAACCCG ATGCTCAGTA 1440

CCCTGACCTC GGCTCTGTCG GGCAAGCTGA ACCCGGATGT GAATCTGGTC GACACCCTCA 1500

ACGGCGGCGA GTACACCGTT TTCGCCCCCA CCAACGCCGC ATTCGACAAG CTGCCGGCGG 1560

CCACTATCGA TCAACTCAAG ACTGACGCCA AGCTGCTCAG CAGCATCCTG ACCTACCACG 1620

TGATAGCCGG CCAGGCGAGT CCGAGCAGGA TCGACGGCAC CCATCAGACC CTGCAAGGTG 1680

CCGACCTGAC GGTGATAGGC GCCCGCGACG ACCTCATGGT CAACAACGCC GGTTTGGTAT 1740

GTGGCGGAGT TCACACCGCC AACGCGACGG TGTACATGAT CGATACGGTG CTGATGCCCC 1800

CGGCACAGTA ACGTTCGGCG CGGTCAAGGC GAGGCAGCCC GTGTAGGCGG TTTGCCTCGC 1860

TCATCCGGCG GCTTCGTGCC GATAGATCAC GTGATATC 1898

CLAIMS

1. The gene (*mpt*83) which encodes the 25 kDa antigen of *Mycobacterium tuberculosis* or a derivative thereof.

2. The nucleotide sequence of the *mpt*83 gene or a derivative thereof.

3. The promoter of the *mpt*83 gene or a derivative thereof.

4. The DNA sequence encoding the glycosylation signal of the *mpt*83 gene or a derivative thereof.

5. The DNA sequence encoding the lipoylation signal of the *mpt*83 gene or a derivative thereof.

6. The DNA sequence encoding the secretion signal of the *mpt*83 gene or derivative thereof.

7. A DNA molecule comprising a promoter as claimed in claim 3 operably linked to a heterologous protein coding sequence.

8. A DNA molecule comprising a sequence encoding a glycosylation signal as claimed in claim 4, or lipoylation signal as claimed in claim 5, or secretion signal as claimed in claim 6 operably linked to a heterologous protein coding sequence.

9. A DNA molecule comprising the following elements operably linked a 5' to 3' direction:

a promoter as claimed in claim 3;

a sequence encoding a glycosylation signal as claimed in claim 4, or lipoylation signal as claimed in claim 5, or secretion signal as claimed in claim 6, and';

a heterologous protein coding sequence.

10. A DNA molecule as claimed in any one of claims 7 to 9 wherein the heterologous protein coding sequence encodes an antigenic protein of a pathogen.

11. A vector or DNA construct including the *mpt*83 gene promoter or a derivative thereof and a sequence encoding the *mpt83* glycosylation signal or a derivative thereof, and one or more restriction endonuclease sites downstream of said *mpt*83 glycosylation sequence such that transcription and translation of a nucleotide sequence inserted into said sites results in a glycosylated transcription and translation product of said nucleotide sequence.

12. A vector or DNA construct including the *mpt*83 gene promoter or a derivative thereof and a sequence encoding the *mpt83* lipoylation signal or a derivative thereof and one or more restriction endonuclease sites downstream of said *mpt*83 lipoylation sequence such that transcription and translation of a nucleotide sequence inserted into said sites results in a lipoylated transcription and translation product of said nucleotide sequence.

13. A vector or DNA construct including the *mpt*83 gene promoter as claimed in claim 3 and its secretion sequence as claimed in claim 6 and one or more restriction endonuclease sites downstream of said *mpt*83 secretion sequence.

14. A vector or DNA construct including the gene, sequence, promoter or DNA molecule as claimed in any one of claims 1 to 10.

15. A host cell transformed or transfected with a vector or DNA construct as claimed in any one of claims 11 to 14.

16. A host cell as claimed in claim 15 which is a strain of *Mycobacterium tuberculosis, Mycobacteria bovis or Mycobacterium bovis* BCG.

17. A method of producing a glycosylated or lipoylated protein, peptide or other product comprising transforming prokaryote or eukaryote cells with a DNA construct or vector as claimed in claim 11 or claim 12.

18. A method of producing a protein, peptide or other product comprising transforming prokaryote or eukaryote cells with a recombinant DNA vector as claimed in claim 13 or claim 14.

19. A method of producing a glycosylated protein as claimed in claim 17 wherein said glycosylation is enhanced by culturing said actinomycete within macrophages or by the use of macrophage factors.

20. The recombinantly produced product of the *mpt*83 gene or a derivative thereof, or portions thereof.

21. A recombinantly glycosylated protein, peptide or other product where said glycosylation is substantially similar to the glycosylation carried out by the *mpt*83 gene.

22. A recombinantly lipoylated protein, peptide or other product where said lipoylation is substantially similar to the lipoylation carried out by the *mpt*83 gene.

23. A protein comprising the following elements linked in a N-terminal to a C-terminal direction:

the glycosylation and or lipoylation signal products of the 25 kDa antigen of *Mycobacterium tuberculosis* or a derivative thereof;

any prokaryote or eukaryote protein.

24. A protein as claimed in claim 23, wherein said prokaryote or eukaryote protein is a heterologous antigenic sequence of a pathogen.

25. A recombinant protein, peptide or other product as claimed in any one of claims 20 to 23 wherein said protein, peptide or other product is a bacterial, viral or other antigen.

26. A vaccine including a protein claimed in claims 24 or 25 or an antigenic epitopic part thereof.

27. A vaccine capable of inducing a serological or cell-mediated immune response diagnostically indistinguishable from a response

induced by wild-type infection with a pathogenic mycobacterium said vaccine comprising a species of microorganism or virus carrying a deletion, insertion or modification of the *mpt*83 gene, or a derivative thereof.

28. An antibody raised against an antigen as claimed in claim 24 or claim 25 or an epitopic part thereof.

29. A vaccine including one or more antibodies as claimed in claim 28.

30. A vaccine including the *mpt*83 gene sequence, a sequence encoding a glycosylation signal as claimed in claim 4, or lipoylation signal as claimed in claim 5, or secretion signal as claimed in claim 6, each being on its own or recombined.

31. A vaccine as claimed in claim 30 wherein said sequences are portions of the full sequences and sufficient to provoke an immune response.

32. A vaccine comprising a pharmaceutically acceptable carrier or diluent and an active ingredient consisting of a host cell as claimed in any one of claims 15 or 16.

33. A host cell according to any one of claims 15 or 16 for use as a vaccine.

34. A host cell as claimed in claim 33 having a native *mpt83* gene or promoter wherein said gene or promoter has been deleted or disrupted such as to render the host cell immunologically distinguishable from untransformed cells.

35. A method for vaccinating a patient against infection by a pathogen, which method comprising administering to the patient a vaccine as claimed in any one of claims 26, 27 or 29 to 32.

36. A diagnostic test kit for testing infection by mycobacteria including the MPT83 protein, or its glycosylation signal, or lipoylation signal, or secretion signal or derivatives thereof or epitopic portions thereof, comprising an antibody or antisera

capable of specifically binding with said protein or signals either unconjugated or conjugated with suitable reagent compounds.

Fig.1.

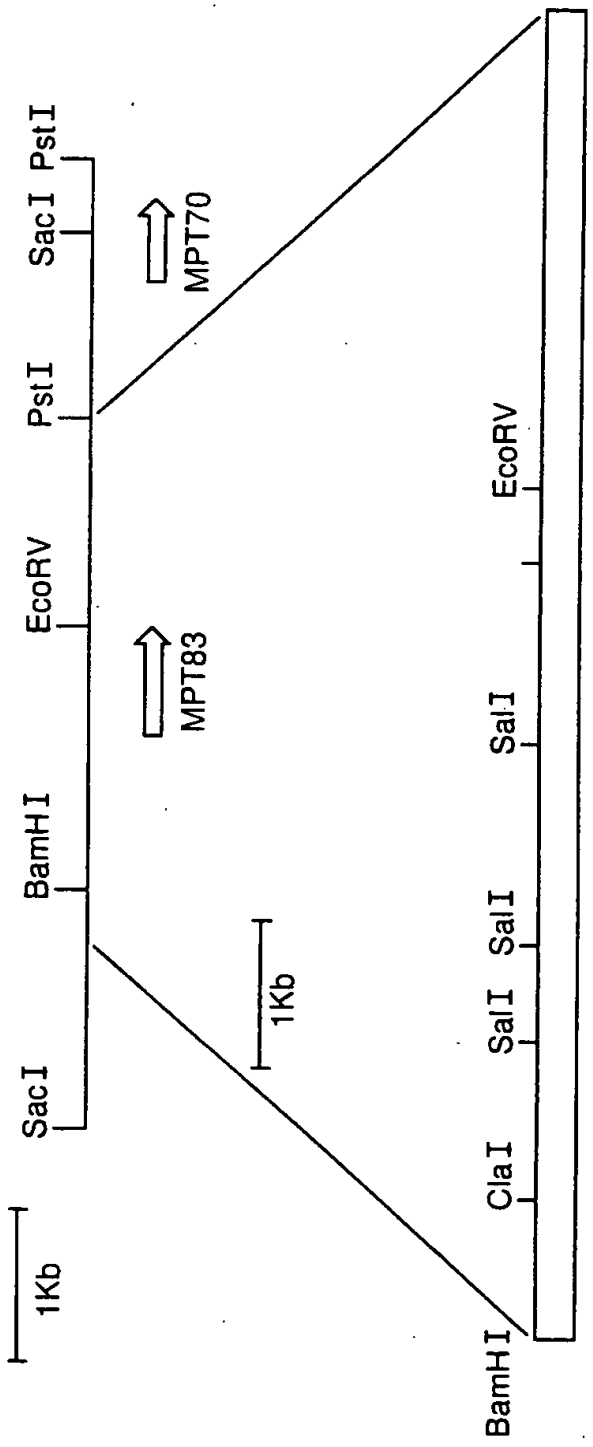

SacI
PstI
MPT70
PstI
EcoRV
MPT83
BamHI
SacI
1Kb
1Kb
EcoRV
SalI
SalI
SalI
ClaI
BamHI

Fig.2A.

```
                    ClaI
  1  ATCAAGCTTA TCGATGACGA GCTGTACCGC GAGGTGAAAS CAAAGSCCGC TCGTTCCGGG
     TAGTTCGAAT AGCTACTGCT CGACATGGCA CTCCACTTTS GTTTCSGGCG AGCAAGGCCC

 61  CGTACCGTGG CCGCGGTTCT TGAAGATGCG GTGCGGCGTG GTCTCAACCC GCCTAAGCCG
     GCATGGCACC GGCGCCAAGA ACTTCTACGC CACGCCGCAC CAGAGTTGGG CGGATTCGGC

121  CAGGCCGCCG GCCGTTATCG AGTCCAGCCG TCGGGTAAGG GCGGBCTBCG GCCCGGTGTC
     GTCCGGCGGC CGGCAATAGC TCAGGTCGGC AGCCCATTCC CGCCVGACGC CGGGCCACAG

181  GATCTATCGT CCAACGCCGC ACTTGBCGAA GCGATGAACG ACGGCGTGTC GGTCGATGCT
     CTAGATAGCA GGTTGCGGCG TGAACVGCTT CGCTACTTGC TGCCGCACAG CCAGCTACGA
                    SalI
241  GTGCGTTGAT GTCGACSTGC TCGTTTACGC GCATCGGGCA GACCTACGGG AGCACGCGGG
     CACGCAACTA CAGCTGSACG AGCAAATGCG CGTAGCCCGT CTGGATGCCC TCGTGCGCCC

301  ACTATCGGGG TTTGCTTGAG CGGCTGGCCA ACGATGACGA GCCGCTGGGT CTACCAGATA
     TGATAGCCCC AAACGAACTC GCCGACCGGT TGCTACTGCT CGGCGACCCA GATGGTCTAT

361  GCGTGCTCGC CGGCTTCATC CGGGTGGTTA CCAACCGCCG CGTCTTCACC GAGCCGACGA
     CGCACGAGCG GCCGAAGTAG GCCCACCAAT GGTTGGCGGC GCAGAAGTGG CTCGGCTGCT
                                 SalI
421  GCCCACAGGA CGCATGGTAG GCAGTCGACG CCCTACTCGC GGXACCCGCA GXCATGCGAC
     CGGGTGTCCT GCGTACCATC CGTCAGCTGC GGGATGAGCG CCXTGGGCGT CXGTACGCTG

481  TTCGGCCTGG XGAGCGCCAC TGGATGGXCT TTCGGCAGTT AGCGTCCCGA TGTTGATGGC
     AAGCCGGACC XCTCGCGGTG ACCTACCXGA AAGCCGTCAA TCGCAGGGCT ACAACTACCG
```

Fig.2A (Cont).

```
541  GAACGGGAAC GACATTGGGG ACGCGCACCT GGGCCGCCTA CGCGYTTAGG AGAACAAMXG
     CTTGCCCTTG CTGTAACCCC TGCGCGTGGA CCCGGCGGAT GCGCRAATCC TCTTGTTKXC

601  CAAACCCTXG TTKGAKXCGX CSCACCCXTC GKXTTTGXCC CYXTTTCCCC TCSRATTTXK
     GTTTGGGAXC AAMCTMXGCX GSGTGGGXAG CMXAAACXGG GRXAAAGGGG AGSYTAAAXM

661  GCTTGAGKYM WCCCMTTGGG GACSGGXYMA ARCMCMYWYT CTTWTYYXGX CCCCAACTTC
     CGAACTCNRK WGGGKAACCC CTGSCCXRKT TYGKGKRWRA GAAWARRXCX GGGGTTGAAG

721  CXAATTXAXT TGXTGTCCAA CCCAAGGAXG GAXGGGGGTT CCAAXXCXCC GXCCGXAAGT
     GXTTAAXTXA ACXACAGGTT GGGTTCCTXC CTXCCCCCAA GGTTXXGXGG CXGGCXTTCA

781  T
     A
```

Fig.2B.

```
   1  GGACCTGATT TGTTCGTGGG TGCGCCACAA TAACGCGGCX CTAAATCCCA CCGGTCCCGG
      CCTGGACTAA ACAAGCACCC ACGCGGTGTT ATTGCGCCGX GATTTAGGGT GGCCAGGGCC

  61  ATYGGTGXCC GGAATGGCGC AAGACCCGGT CGCTACCGCG GCTTCCAACA ACCCGATGCT
      TARCCACXGG CCTTACCGCG TTCTGGGCCA GCGATGGCGC CGAAGGTTGT TGGGCTACGA
                                                          SalI
 121  CAGTACCCTG ACCTCGGCTC TGTCGGGCAA GCTGAACCCG GATGTGAATC TGGTCGACAC
      GTCATGGGAC TGGAGCCGAG ACAGCCCGTT CGACTTGGGC CTACACTTAG ACCAGCTGTG

 181  CCTCAACGGC GGCGAGTACA CCGTTTTCGC CCCCACCAAC GCCGCATTCG ACAAGCTGCC
      GGAGTTGCCG CCGCTCATGT GGCAAAAGCG GGGGTGGTTG CGGCGTAAGC TGTTCGACGG

 241  GGCGGCCACT ATCGATCAAC TCAAGACTGA CGCCAAGCTG CTCAGCAGCA TCCTGACCTA
      CCGCCGGTGA TAGCTAGTTG AGTTCTGACT GCGGTTCGAC GAGTCGTCGT AGGACTGGAT

 301  CCACGTGATA GCCGGCCAGG CGAGTCCGAG CAGGATCGAC GGCACCCATC AGACCCTGCA
      GGTGCACTAT CGGCCGGTCC GCTCAGGCTC GTCCTAGCTG CCGTGGGTAG TCTGGGACGT

 361  AGGTGCCGAC CTGACGGTGA TAGGCGCCCG CGACGACCTC ATGGTCAACA ACGCCGGTTT
      TCCACGGCTG GACTGCCACT ATCCGCGGGC GCTGCTGGAG TACCAGTTGT TGCGGCCAAA

 421  GGTATGTGGC GGAGTTCACA CCGCCAACGC GACGGTGTAC ATGATCGATA CGGTGCTGAT
      CCATACACCG CCTCAAGTGT GGCGGTTGCG CTGCCACATG TACTAGCTAT GCCACGACTA

 481  GCCCCCGGCA CAGTAACGTT CGGCGCGGTC AAGGCGAGGC AGCCCGTGTA GGCGGTTTGC
      CGGGGGCCGT GTCATTGCAA GCCGCGCCAG TTCCGCTCCG TCGGGCACAT CCGCCAAACG
                                                  EcoRV
 541  CTCGCTCATC CGGCGGCTTC GTGCCGATAG ATCACGTGAT ATC
      GAGCGAGTAG GCCGCCGAAG CACGGCTATC TAGTGCACTA TAG
```

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/GB 96/02015

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/31 C12N15/76 C12N15/62 C12N1/21 C12N7/01
C07K14/35 C07K16/12 A61K39/04 A61K39/40 G01N33/569
C12Q1/04 //(C12N1/21,C12R1:32)
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	SCAND. J. IMMUNOL., vol. 43, no. 5, May 1996, pages 490-499, XP000611375 HEWINSON ET AL.: "Molecular characterization of MPT83: a seroreactive antigen of Mycobacterium tuberculosis with homology to MPT70" see the whole document ---	1-6, 14-16, 20-23, 28-30, 32,33, 35,36
A	SCAND. J. IMMUNOL., vol. 42, no. 1, July 1995, pages 46-51, XP000611303 HARBOE ET AL.: "Homology between the MPB70 and MPB83 proteins of Mycobacterium bovis BCG" see the whole document --- -/--	1-36

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 November 1996	03. 12. 96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Gac, G

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SCAND. J. IMMUNOL., vol. 41, no. 3, March 1995, pages 281-287, XP000611440 MATSUMOTO ET AL.: "Cloning and sequencing of a unique antigen MPT70 from Mycobacterium tuberculosis H37Rv and expression in BCG using E. coli-Mycobactria shuttle vector" see the whole document ---	1-36
A	INFECT. IMMUN., vol. 59, no. 3, 1991, pages 800-807, XP000178825 FIFIS ET AL.: "Purification and characteization of major antigens from a Mycobacterium bovis culture filtrate" cited in the application see the whole document ---	1-36
A	WO 90 10701 A (COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION) 20 September 1990 see the whole document -----	1-36

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 96/02015

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 35
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: Although claim 35 is directed to a method of treatment of the human/animal body (Art.17.2a)i) and Rule 39.2.iv PCT, the search has been carried out and based on the alleged effects of the compound/ composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

In' ional Application No
PCT/GB 96/02015

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9010701	20-09-90	AU-B-	627011	13-08-92
		AU-A-	5262690	09-10-90
		EP-A-	0486495	27-05-92